Docket No. NEB-135

COPY

IN THE UNITED STATES PATENT OFFICE AND TRADEMARK OFFICE
APPLICATION FOR UNITED STATES LETTERS PATENT

INVENTOR(S): Jay Wayne
Shuang-yong Xu

TITLE: METHOD FOR CONSTRUCTION OF *THERMUS-E. COLI* SHUTTLE VECTORS AND IDENTIFICATION OF TWO *THERMUS* PLASMID REPLICATION ORIGINS

ATTORNEY: Gregory D. Williams
General Counsel
NEW ENGLAND BIOLABS, INC.
32 Tozer Road
Beverly, Massachusetts 01915
(978) 927-5054; Ext. 292

EXPRESS MAILING LABEL NO.: IB442856348US

METHOD FOR CONSTRUCTION OF *THERMUS-E. COLI* SHUTTLE VECTORS AND IDENTIFICATION OF TWO *THERMUS* PLASMID REPLICATION ORIGINS

BACKGROUND OF THE INVENTION

The present invention relates to recombinant DNA molecules encoding plasmid DNA replication origins in *Thermus*, as well as to shuttle vectors which contain the same.

Many species of bacteria contain small circular extrachromosomal genetic elements, known as plasmids. Plasmids have been found in a number of bacteria which live in extreme environments, including the thermophiles, which live at high temperatures of more than 55°C (Munster et al., *Appl. Environ. Microbiol.* 50:1325-1327 (1985); Kristjansson and Stetter, in 'Thermophilic Bacteria', Kristjansson, ed., p. 1-18 (1992)). However, most thermophile plasmids remain 'cryptic' in that functional genes have not been isolated from them, hence leaving their functional significance speculative (Hishinuma et al., *J. Gen. Microbiol.* 104:193-199 (1978); Eberhard et al., *Plasmid* 6:1-6 (1981); Vásquez et al., *FEBS Lett.* 158:339-342 (1983)). Common genes found in plasmids include those encoding plasmid replication and cellular maintenance, antibiotic resistance, bacteriocin production, sex determination, and other cellular functions (Kornberg and Baker, 'DNA Replication', 2nd ed. (1991)).

It is often particularly difficult to cultivate thermophilic bacteria within the laboratory. They require high temperatures and often-unknown environmental conditions for acceptable growth (Kristjansson and Stetter, in 'Thermophilic Bacteria', Kristjansson, ed., p. 1-18 (1992)). However, with the advent of genetic engineering, it is now possible to clone genes from thermophiles into more easily cultivatable laboratory organisms, such as *E. coli* (Kristjansson, *Trends Biotech.* 7:349-353 (1989); Coolbear et al., *Adv. Biochem. Eng. Biotech.* 45:57-98 (1992)). The expression of such genes can be finely controlled within *E. coli*.

A *Thermus-E. coli* shuttle vector would be desirable if one needs to have the convenience of cloning in *E. coli*, isolation of DNA from *E. coli* for further manipulations and subsequently gene selection and expression in *Thermus*. Such *Thermus-E. coli* shuttle vectors could be used to screen, select and express thermostable proteins in *Thermus*. Using these vectors, a gene could, for example, be mutated within a mesophile, transferred to a thermophile, and then its encoded protein selected for increased thermostability. In this way, mesophile-thermophile shuttle-vectors can be used to conduct directed evolution, or protein engineering, on desirable gene products.

There is commercial incentive to produce thermostable proteins which are usually more thermostable in denaturing conditions then mesophilic counterparts (Wiegel and

Ljungdahl, *CRC Crit. Rev. Biotech.* 3:39-108 (1984); Kristjansson, *Trends Biotech.* 7:349-353 (1989); Coolbear et al., *Adv. Biochem. Eng. Biotech.* 45:57-98 (1992)). These thermostable enzymes can also be used in a variety of assays, such as PCR, restriction enzyme-mediated PCR, thermo-cycle DNA sequencing and strand-displacement amplification, in which high temperatures are desirable. The shuttle vectors of the present invention should facilitate production of such thermostable proteins.

SUMMARY OF THE INVENTION

The present invention relates to recombinant DNA molecules encoding plasmid DNA replication origins in *Thermus*, as well as to shuttle vectors which contain the same.

Mesophile-thermophile shuttle vectors require origins of replication (*ori*s) to be genetically maintained and transferred within each bacterial species. To construct appropriate mesophile-thermophile shuttle-vectors, restriction digested thermophile plasmid DNA fragments were ligated into the mesophilic vector pUC19-Km^R (the thermostable Km^R marker can be selected at 50°-65°C). Plasmid pUC19 uses the ColEI *ori* to replicate within *E. coli*, and does not replicate within the plasmid-accepting thermophile *Thermus thermophilus* HB27 or HB27 Pro^- (Koyama et al., *J. Bacteriol.* 166:338-340 (1986)). We reasoned that the introduction of plasmid DNA

from related *Thermus* species, which contained a complete thermophilic *ori*, would confer plasmid replication within HB27.

The thermophilic eubacterium *Thermus* species YS45 (Raven et al., *Nucl. Acids Res.* 21:4397 (1993)) contains two cryptic plasmids, and grows between 55°C and 70°C. These two *Thermus* plasmids were named pTsp45S and pTsp45L. These plasmids were digested with a variety of restriction endonucleases to produce fragments that can be cloned into pUC19-derived vectors. A pUC19-derived plasmid with a 4.2-kb *Xba*I fragment of the small plasmid (pTsp45S, 5.8 kb) of YS45 replicated within HB27. Therefore this *Xba*I fragment must contain a thermophilic *ori*. Subsequent deletion analysis revealed that only 2.3 kb (an *Nhe*I fragment) within the 4.2 kb was necessary for thermophilic plasmid replication, and that it encodes a replication protein (RepT). The *repT* gene encodes the 341 amino acid protein, RepT, with predicted molecular mass of 38.2 kDa.

A second *Thermus* plasmid replication origin from pTsp45L was defined within a 9 kb *Sph*I fragment. This fragment encodes a gene (*parA*) for plasmid replication and partition. It also contains direct repeats of 5' RRCTTTTYYY 3' (SEQ ID NO:1), 5' RRYTTTG 3' (SERQ ID NO:2), and an inverted repeat of

5' TTAACCTTTTTTCAAGAAAAAGAGATAA 3' (SEQ ID NO:3)
3' AATTGGAAAAAAGTT CTTTTTCTCTATT 5'
(COMPLEMENT OF SEQ ID NO:3)

The direct repeats and inverted repeats are important for pTsp45L plasmid replication. Deletion of these repeats abolished replication activity in *Thermus*.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is the DNA sequence (SEQ ID NO:4) of *repT* gene from pTsp45S and its encoded amino acid sequence.

Figure 2 is the promotor sequence (SEQ ID NO:5) upstream of *repT* gene.

Figure 3 is the entire DNA sequence (SEQ ID NO:6) of *Thermus* plasmid pTsp45S.

Figure 4 illustrates the genetic organization of *Thermus* plasmid pTsp45S. The gene *repT* encodes RepT for plasmid replication.

Figure 5 is the *parA* DNA sequence from pTsp45L and the encoded amino acid sequence (SEQ ID NO:7).

Figure 6 is the seven open reading frames encoded by pTsp45L. Frames a, b, and c are encoded by the top strand. Frames d, e, and f are encoded by the bottom strand.

Figure 7 is the entire DNA sequence of *Thermus* plasmid pTsp45L (SEQ ID NO:8).

DETAILED DESCRIPTION OF THE INVENTION

The method described herein by which a *Thermus* plasmid replication origin is preferably cloned and selected comprises the following steps:

1. The plasmid DNA of a target host, such as *Thermus* species YS45 plasmid pTsp45S and pTsp45L, is purified.

2. The plasmid DNA is digested with appropriate restriction endonucleases; for *Thermus* species YS45, *Hin*dIII, *Kpn*I, *Pst*I, *Sph*I, and *Xba*I are used to generate 1 to 12 kb restriction fragments. This map is used to orient and localize genes within the plasmid.

3. The digested plasmid DNA is then ligated into similarly cleaved/CIP treated vectors such as pUC-EKR or pUC-EKF (Ap^R at 37°C, Km^R at 50-65°C) cloning vectors. The ligated DNA is used to transform an appropriate host, e.g., a $HsdR^-$, $McrBC^-$, Mrr^- strain, such as *E. coli* strain RR1. The DNA/cell mixtures are then plated on ampicillin selective media to grow only transformed cells to form primary restriction libraries, such as *Hin*dIII, *Kpn*I, *Pst*I, *Sph*I, and *Xba*I DNA libraries for Thermus species YS45.

4. The recombinant plasmids are purified to form the primary plasmid library that might contain thermophilic

plasmid origins. Plasmids are digested *in vitro* with a variety of endonucleases to confirm DNA inserts.

5. The plasmid DNA libraries are used to transform an appropriate thermophilic host cell such as *Thermus thermophilus* HB27 (Pro^-) cells and transformants are selected on Km plates at 60°-65°C for 48 hours.

6. Individual Km^R transformants are amplified in small culture at 65°C and plasmid DNA is isolated from the overnight cell culture. The plasmid DNA is then digested with an appropriate restriction endonuclease (e.g., *Hin*dIII, *Kpn*I, *Pst*I, *Sph*I, or *Xba*I) to cut out the *Thermus* DNA insert.

7. One clone from the *Xba*I library described above contained a 4.2 kb *Thermus* DNA which replicates in both *Thermus* and *E. coli*. The 4.2 kb insert DNA of the recombinant pUC-EKF clone was sequenced. To facilitate sequencing, the insert DNA wass further sub-cloned within pUC19 based upon preliminary sequence and mapping. The sequenced DNA was then assembled to match that of the thermophilic plasmid map. The remaining DNA fragments from pTsp45S were also cloned and sequenced. In this way, the thermophilic plasmid (pTsp45S) was completely sequenced.

8. To reduce the size of the *Thermus* replication origin, the 4.2 kb *Xba*I fragment was further digested with restriction enzymes and subcloned into pUC-EKF or pUC-EKR.

One recombinant plasmid contained a 2.3 kb *Nhe*I fragment that replicates in *Thermus* and *E. coli*. This plasmid pUC-EKF-Tsp3 is a *Thermus-E. coli* shuttle vector.

9. One open reading frame of 1026 bp encoding a 341-amino acid protein was found within the *Thermus* origin. Deletion of 234 bp (78 amino acid residues) within this gene abolished the *Thermus* replication function. Insertion of stop codons within this gene causes premature termination and negates the *Thermus* transformation. Therefore it was determined that this gene (*repT*) is required for plasmid replication in *Thermus* HB27 (Pro^-) cells.

10. Two *Thermus* promoters were found upstream of the *repT* gene that are important for *repT* expression.

11. Plasmid pTsp45L (a mixture of pTsp45L and pTsp45S) was digested with *Hin*dIII, *Kpn*I, *Pst*I, *Sph*I, or *Xba*I. The digested DNA fragments were cloned into pUC-EKR vector to produce *Thermus* DNA libraries for subsequent selection of *Thermus* plasmid replication origin(s).

12. Approximately 450 Ap^R transformants were derived from pUC-EKR + *Hin*dIII fragments, + *Kpn*I fragments, + *Pst*I fragments, + *Sph*I fragments, and + *Xba*I fragments, respectively. pUC-EKR plasmids with *Hin*dIII, *Kpn*I, *Pst*I, *Sph*I, or *Xba*I fragment inserts were amplified in *E.coli*.

13. The DNA libraries were used to transform *Thermus thermophilus* HB27 (Pro^-). Transformants were plated on Km plates and incubated at 60°C for two days. Plasmid DNA was extracted from seventeen Km^R transformants and digested with *Xba*I, *Pst*I, or *Sph*I. Restriction mapping and Southern blot analysis were carried out.

14. The 9 kb *Sph*I *Thermus* origin insert and the 12 kb *Thermus* origin insert were from pTsp45L. The entire pTsp45L plasmid can be separated into two *Sph*I fragments, 3 kb and 9 kb respectively. The 9 kb *Sph*I fragment contains the functional *Thermus* replication origin. The inserts were sequenced by using pUC19 universal forward and reverse primers and by primer walking. Plasmid pTsp45L is 11958 bp, encoding 7 possible genes.

15. Orf3 is most likely the candidate for pTsp45L replication protein, because it has homolgy to RepA protein of *Agrobacterium* plasmid pTiB6S3, replication protein of *Agrobacterium* plasmid pRiA4b, plasmid partition protein of *Borrelia*, partition protein of *Frankia*, RepA protein of *Rhizobium*, and DNA partition protein ParA of *Caulobacter*. Orf2 may be an accessary protein for pTsp45L plasmid replication. Orf3 was renamed as *parA* gene.

16. There are direct repeats and inverted repeats in the 9 kb *Sph*I fragment containing the functional replication origin. The direct repeats I are:

5' GGCTTTTCTT 3' (SEQ ID NO:9)
5' AACTTTTCCC 3' (SEQ ID NO:10)
5' GACTTTTTTC 3' (SEQ ID NO:11)
consensus 5' RRCTTTTYYY 3' (SEQ ID NO:1)

The direct repeats II are:

5' AACTTTG 3' (SEQ ID NO:12)
5' AGTTTTG 3' (SEQ ID NO:13)
5' GATTTTG 3' (SEQ ID NO:14)
5' AACTTTG 3' (SEQ ID NO:15)
consensus 5' RRYTTTG 3' (SEQ ID NO:2)

The inverted repeat is:

5' TTAACCTTTTTTCAAGAAAAAGAGATAA 3' (SEQ ID NO:3)
3' AATTGGAAAAAAGTT CTTTTTCTCTATT 5'
(COMPLEMENT OF SEQ ID NO:3)

(underlined bases are inverted repeat).

Deletion of these repeats in a *Hin*dIII fragment abolished DNA replication in *Thermus*.

Any *Thermus* plasmid DNA, *Thermus* viral DNA, or genomic DNA can be digested with restriction enzymes to generate 2 - 20 kb fragments. The restriction fragments can be ligated with similarly-cut pUC-EKF or pUC-EKR and transformed into *Thermus* cells and selected for Km^R transformants. Alternatively, DNA can be extracted from

environmental samples, such as water from hot springs and soil sediment from hot springs, digested with restriction enzymes, ligated into similarly-cut pUC-EKF or pUC-EKR and transformed into *Thermus* cells and selected for Km^R transformants. Because of the small amount of DNA from environmental samples, one can transfer such DNA into *E. coli* first to amplify the DNA library and then transform such DNA into *Thermus.*

The following Examples are given to illustrate embodiments of the present invention, as it is presently preferred to practice. It will be understood that these Examples are illustrative, and that the invention is not to be considered as restricted thereto except as indicated in the appended claims.

The references cited above and below are herein incorporated by reference.

EXAMPLE I

1. Cloning of a replication origin from a *Thermus* plasmid pTsp45S native to *Thermus* species YS45.

Thermus species YS45 (Raven et al., *Nucl. Acids Res.* 21:4397 (1993) obtained from R.A.D. Williams of Queen Mary and Westerfield College, University of London) can be grown in modified *Thermus thermophilus* liquid media (Oshima and

Imahori, *J. Sys. Bacteriol.* 24:102-112 (1974)) consisting of 0.5% tryptone (DIFCO Laboratories; Detroit, Michigan), 0.4% yeast extract (DIFCO Laboratories; Detroit, Michigan), 0.2% NaCl at pH 7.5. Cells are plated in this media with 3% agar. Plated colonies are distinguishable after two days incubation at 55°-70°C. Individual colonies form dense liquid overnight cultures (3-10 ml) at 55°-70°C in a shaking waterbath. One-ml aliquots of overnight cultures are pelleted and stored at –20°C for up to one month without loss of viability. Overnight cultures are also stably maintained in media with 25% glycerol at -70°C.

Ten ml of 70°C overnight YS45 culture is diluted 1:1000 in 500 ml of *Thermus* media, and grown overnight at 70°C to generate plasmid DNA. Plasmid DNA is prepared via the Qiagen mid-prep protocol (Qiagen, Inc.; Studio City, California) with the addition of 2 mg lysozyme per ml. Lysis is very inefficient without the presence of lysozyme in the first resuspension buffer (Oshima and Imahori, *J. Sys. Bacteriol.* 24:102-112 (1974)). Routinely, between 50-150 μg of plasmid DNA is obtained from 500 ml of overnight YS45 culture.

YS45 contains two plasmids of 5.8 kb (pTsp45S) and approximately 12 kb (pTsp45L) (Wayne and Xu, *Gene* 195:321-328 (1997)). Each plasmid contains a single *Pst*I site useful for linearizing and visualizing the plasmids on agarose gels. Plasmid pTsp45S also contains two *Xba*I sites that generate

4.2 and 1.6-kb fragments. This plasmid is extensively mapped and cloned into pUC19 as three fragments: 4.2-kb *Xba*I-*Xba*I, 0.7-kb *Xba*I-*Pst*I, and 0.9-kb *Pst*I-*Xba*I. The 4.2-kb fragment is then further mapped and sub-cloned into pUC19 as six smaller fragments: 0.4-kb *Xba*I-*Hin*dIII, 1.1-kb *Hin*dIII-*Hin*dIII, 0.7-kb *Hin*dIII-*Hin*dIII, 0.5-kb *Hin*dIII-*Sca*I, 1.0-kb *Sca*I-*Sca*I, and 0.5-kb *Sca*I-*Xba*I. Cloning was accomplished by isolating digested fragments from agarose gels and combining them with compatibly cut pUC19 by standard methods (Sambrook et al., 'Molecular Cloning A Laboratory Manual', 2nd ed. (1989)).

The clones are sequenced using universal and reverse M13/pUC primers (New England Biolabs, Inc.; Beverly, Massachusetts). Preliminary sequencing was used to generate 12 additional primers (synthesized at New England Biolabs, Inc.; Beverly, Massachusetts) to refine and correct sequencing errors. The primers (shown as top and bottom strand pairs) are:

5'-GGTTCCATAAGGCGGGTCAATATAG-3' (SEQ ID NO:16);
5'-CTATATTGACCCGCCTTATGGAACC-3' (SEQ ID NO:17);
5'-GT GGGGTGGGCTGATCAAGAATCTCCT-3' (SEQ ID NO:18);
5'-AGGAGATTCTTGATCAGCCCACCCCAC-3' (SEQ ID NO:19);
5'-TCACCCACAACCCTCACGCACTCCAA-3' (SEQ ID NO:20);
5'-TTGGAGTGCGTGAGGGTTGTGGGTGA-3' (SEQ ID NO:21);
5'-AGATGTAGTCGTCCAGGGTGAGCCTG-3' (SEQ ID NO:22);
5'-CAGGCTCACCCTGGACGACTACATCT-3' (SEQ ID NO:23);

5'-TTGGTATGTAAAGCCCTTCGCGAGG-3' (SEQ ID NO:24);
5'-CCTCGCGAAGGGCTTTACATACCAA-3' (SEQ ID NO:25);
5'-TAGTGGCATCGGTGTTGTCGTGGGT-3' (SEQ ID NO:26); and
5'-ACCCACGACAACACCGATGCCACTA-3' (SEQ ID NO:27)

(underlined bases are in pTsp45s, but were not originally synthesized in these primers).

2. Characteristics of a thermophilic plasmid *ori*

The 2.3-kb *Nhe*I-bounded thermophilic *ori* is 57% G + C. The 5.8-kb *Thermus* plasmid pTsp45S is 54% G + C, and there are no other published reports of the G + C content in its natural host, YS45. There are no significant AT-rich regions within the sequenced *ori*.

The thermophilic *ori* contained one significant ORF of 1026 bp, beginning with GTG and ending with TGA (Figure 1). The ORF's 341 amino acid could encode a protein with a predicted molecular weight of 38.2 kDa. Centered 10 bp 5' of this ORF is a putative RBS, GGAGG (Hartmann and Erdmann, *J. Bacteriol.*, 171:2933-2941 (1989)). Further upstream, two possible promoter regions (-10 TATTTT, -35, TTGCCA, 17 bp spacing; or -10 TAGGGT, -35 TTGCCC, 18 bp spacing) were found (Figure 2) with significant homology to the *Thermus* consensus promoter (Maseda and Hoshino *FEMS Microbiol. Lett.* 128:127-134 (1985)). Database searches (FASTA, BLAST) did

not reveal any significant homologies to the predicted protein, or to other possible reading frames.

To test the importance of this ORF in the thermophilic replication, a significant portion of it was deleted. Briefly, pUC-EKF-Tsp3 was digested with *Nru*I + *Psh*AI, removing 234 bp or 78 aa within the ORF. The linearized plasmid was self-ligated, generating pUC-EKF-Tsp3-ΔNP(7.5 kb), then amplified in *E. coli* and used to transform HB27. No pUC-EKF-Tsp3-ΔNP(7.5 kb) Km^R transformants were found. It was concluded that 234 bp deletion within the *repT* gene abolished the replication function. Similarly, the addition of an *Xba*I amber stop linker (CTAGTCTAGACTAG (SEQ ID NO:28)) at either the *Nru*I or *Psh*AI site of pUC-EKF-Tsp3 negated thermophilic transformation. This indicated that the *repT* within the *Nhe*I fragment was necessary for replication in the thermophile. We suggest that this ORF of pTsp45S is a novel replication protein (RepT) needed for thermophilic plasmid replication. In addition, analysis of this thermophilic *ori* revealed two sequences with significant homology to highly conserved DnaA boxes. Although not yet described in *Thermus*, DnaA boxes are required for binding of a DnaA protein, and for subsequent replication of some plasmids (McMacken, et al., DNA Replication (Chapter 39), pages 586-587 in *Escherichia coli* and *Salmonella typhimmarium*, American Society for Microbiology, Washington, DC) . Both putative DnaA boxes (TTATCACCC (SEQ ID NO:29), TTATCCGAG (SEQ ID NO:30)) of pUC-EKF-Tsp3 lie within the 3' end of *repT*, and are not within

the region deleted in pUC-EKF-Tsp3-ΔNP. Plasmid copy number might be regulated by the relationship between binding of a DnaA homologue at these sites, and transcription of *repT*.

A sample of ER2688[pUC-EKF-Tsp3] has been deposited under the terms and conditions of the Budapest Treaty at the American Type Culture Collection on June 22, 1998, 1998 and received ATCC Accession No. 98793.

EXAMPLE II

Thermus YS45 strain contains two plasmids of 5.8 kb (pTsp45S) and approximately 12 kb (pTsp45L) (Wayne and Xu, *Gene* 195:321-328 (1997)). Each plasmid contains a single *Pst*I site useful for linearizing and visualizing the plasmids on agarose gels. The two plasmid mixture was digested with *Hin*dIII, *Kpn*I, *Pst*I, *Sph*I, or *Xba*I. The digested DNA fragments were cloned into pUC-EKR vector to produce *Thermus* DNA libraries and for subsequent selection of *Thermus* plasmid replication origin(s). Approximately 100, 100, 100, 100, and 50 Ap^R transformants were derived from pUC-EKR + *Hin*dIII fragments, + *Kpn*I fragments, + *Pst*I fragments, + *Sph*I fragments, and + *Xba*I fragments, respectively. Plasmids pUC-EKR with *Hin*dIII, *Kpn*I, *Pst*I, *Sph*I, or *Xba*I fragment inserts were amplified in *E. coli* and the DNA libraries were used to transform *Thermus thermophilus* HB27 (Pro^-). Transformants were plated on Km plates and incubated at 60°C for two days. Plasmid DNA was extracted from seventeen Km^R

transformants and digested with *Xba*I, *Pst*I, or *Sph*I. Restriction mapping and Southern blot analysis indicated that the 4.2 kb *Xba*I fragment *Thermus* origin insert was from pTsp45S, the 9 kb *Sph*I *Thermus* origin insert and the 12 kb *Thermus* origin insert were from pTsp45L. It was concluded that the entire pTsp45L plasmid can be separated into two *Sph*I fragments, 3 kb and 9 kb respectively. The 9 kb *Sph*I fragment contains the functional *Thermus* replication origin. The two *Sph*I fragments were sequenced by subcloning of one *Bam*HI fragment (1.4 kb), one *Hin*dIII fragment (1.9 kb), one *Sph*I fragment (3 kb), two *Kpn*I fragments (2.5 kb, 0.6 kb), three *Sac*I fragments (4.3 kb, 1.9 kb, 1.3 kb), and multiple *Sma*I fragments into pUC19. The inserts were sequenced by using pUC19 universal forward and reverse primers and by primer walking. Plasmid pTsp45L is 11958 bp, encoding 7 possible genes. These seven genes are named orf1 through orf7 (Figure 6). Orf1 amino acid sequence has weak similarity to transposases. Orf3 amino acid sequence has similarity to DNA replication protein RepA and DNA partition protein ParA. Orf4 amino acid sequence has similarity to serine carboxy peptidase III. Orf5 amino acid sequence has similarity to UvrB protein. Orf2, orf6, and orf7 amino acid sequences have no homologs to proteins in Genbank. The 3 kb *Sph*I fragment contains orf5 C-terminus portion, orf6 and orf7. Deletion of this 3 kb did not affect pTsp45L plasmid origin of replication. It was concluded that orfs 5, 6, and 7 are not required for plasmid replication. The 9 kb *Sph*I fragment contains the functional replication origin, which contains orf1, 2, 3, 4 and

a portion of orf5. Orf1 and orf4 have homology to transposases and proteases, respectively. It was concluded that orf1 and orf4 are unlikely involved in DNA replication and that orf3 is most likely the candidate for pTsp45L replication protein, because it has homolgy to RepA protein of *Agrobacterium* plasmid pTiB6S3, replication protein of *Agrobacterium* plasmid pRiA4b, plasmid partition protein of *Borrelia*, partition protein of *Frankia*, RepA protein of *Rhizobium*, and DNA partition protein ParA of *Caulobacter*. Orf2 may be an accessary protein for pTsp45L plasmid replication. Orf3 (coordinate 5876 to 6478) was renamed as *parA* gene. The DNA sequence and amino acid sequence of *parA* is shown in Figure 5. The location, direction, and organization of the seven open reading frames in pTsp45L are shown in Figure 6.

There are direct repeats and inverted repeats in the 9 kb *Sph*I fragment containing the functional replication origin. The direct repeats I are:

5' GGCTTTTCTT 3' (SEQ ID NO:9)
5' AACTTTTCCC 3' (SEQ ID NO:10)
5' GACTTTTTTC 3' (SEQ ID NO:11)
consensus 5' RRCTTTTYYY 3' (SEQ ID NO:1)

The direct repeats II are:

5' AACTTTG 3' (SEQ ID NO:12)
5' AGTTTTG 3' (SEQ ID NO:13)
5' GATTTTG 3' :(SEQ ID NO:14)
5' AACTTTG 3' (SEQ ID NO:15)
consensus 5' RRYTTTG 3' :(SEQ ID NO:2)

The inverted repeat is:

5' TTAACCTTTTTTCAAGAAAAAGAGATAA 3' (SEQ ID NO:3)
3' AATTGGAAAAAAGTT CTTTTTCTCTATT 5'
(COMPLEMENT OF SEQ ID NO:3)

(underlined bases are inverted repeats).

The repeats and inverted repeats are important for pTsp45L origin of replication, because deletion of these repeats in a *Hin*dIII fragment abolished DNA replication in *Thermus*. The DNA sequence of pTsp45L is shown in Figure 7. The *Thermus-E. coli* shuttle vector containing pTsp45L DNA replication origin was named as pUC-EKR-Tsp45L9Kb.

A sample of ER2688[pUC-EKR-Tsp45L9kb] has been deposited under the terms and conditions of the Budapest Treaty at the American Type Culture Collection on June 22, 1998, and received ATCC Accession No. 98794.

EXAMPLE III

Thermus strain YS45 (Raven, et al., *Nucl. Acids Res.* 21:4397 (1993) obtained from R.A.D. Williams of Queen Mary and Westerfield College, University of London) also harbors a plasmid. Plasmid DNA was extracted from *Thermus* species YS45 by midi Qiagen column. The plasmid DNA was cleaved with *Hin*dIII, *Kpn*I, *Pst*I, *Sph*I, or *Xba*I. The digested DNA fragments were cloned into pUC-EKR vector to produce *Thermus* DNA libraries and for subsequent selection of *Thermus* plasmid replication origin(s). Approximately 50 to

300 Ap^R *E. coli* transformants were derived from pUC-EKR + *Hin*dIII fragments, + *Kpn*I fragments, + *Pst*I fragments, + *Sph*I fragments, and + *Xba*I fragments, respectively. Plasmids pUC-EKR with *Hin*dIII, *Kpn*I, *Pst*I, *Sph*I, and *Xba*I fragment inserts were amplified in *E. coli* and the DNA libraries were used to transform *Thermus thermophilus* HB27 (Pro^-). Transformants were plated on Km plates and incubated at 60°C for two days. *Thermus* transformants were found in *Hin*dIII and *Pst*I DNA libraries. Plasmid DNA was extracted from seventeen Km^R *Thermus* transformants and digested with *Hin*dIII or *Pst*I. It was found that the functional Tse plasmid replication origin was contained in a ~7 kb *Hin*dIII or *Pst*I fragment. The shuttle vector was named pUC-EKR-Tse7Kb.

EXAMPLE IV

Thermus cells can be grown in modified *Thermus thermophilus* liquid media (Oshima and Imahori, *J. Sys. Bacteriol.* 24:102-112 (1974)) consisting of 0.5% tryptone (DIFCO Laboratories; Detroit, Michigan), 0.4% yeast extract (DIFCO Laboratories; Detroit, Michigan), 0.2% NaCl at pH 7.5. *Thermus* cells can also be cultured in 4 to 10-fold diluted rich both at 50°-75°C. Ten ml of overnight cell culture is diluted 1:1000 in 500 ml of *Thermus* media, and grown overnight at 50°-75°C to generate plasmid DNA. Plasmid DNA can be prepared via the Qiagen midi/maxi-prep protocol (Qiagen, Inc.; Studi City, California) with the addition of 2 mg lysozyme per ml or any other plasmid preparation method such as alkaline

lysis or boiling methods. The purified plasmid DNA can be digested with restriction enzymes to produce DNA fragments of 2 to 20 kb. The plasmid DNA can also be sonicated to produce blunt end framgents and be made into sticky ends by addition of deoxynucleotides by terminal nucleotide transferase. The DNA fragments can be cloned into pUC-EKF or pUC19-EKR and the ligated DNA can be used for thermophilic transformation into *Thermus* cells. Transformants can be selected by plating cells on Km plates. Any Km^R transformants should contain *Thermus* plasmid replication origin. The origin can be further subcloned and sequenced. A minimal replication origin can be defined by subcloning smaller DNA fragments into pUC-EKF or pUC19-EKR and the resulting plasmid DNA can be used for thermophilic transformation.

Alternatively, plasmid DNA, *Thermus* viral DNA or genomic DNA can be extracted from environmental samples such as water from hot springs and soil sediment from hot springs and digested with restriction enzymes and ligated into similarly-cut pUC-EKF or pUC-EKR. The ligated DNA can be transformed into *Thermus* cells and select for Km^R transformants. Because of the small amount of DNA from environment samples, one can transfer DNA into *E. coli* first to amplify DNA library and then transform into *Thermus*. The thermophilic replication origin can be further subcloned and sequenced. A minimal replication origin can defined by subcloning smaller DNA fragments into pUC-EKF or pUC19-EKR

and the resulting plasmid DNA can be used for thermophilic transformation.